ENZYME KINETICS

From diastase to multi-enzyme systems

ARTHUR R. SCHULZ
Indiana University School of Medicine

CAMBRIDGE
UNIVERSITY PRESS

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1994

Library of Congress Cataloging-in-Publication Data

Schulz, Arthur R.
Enzyme kinetics from diastase to multi-enzyme systems / Arthur R. Schulz
p. cm.
ISBN 0-521-44500-0. – ISBN 0-521-44950-2 (pbk.)
1. Enzyme kinetics. I. Title.
QP601.3.S38 1994 94-28309
574.19'25 – dc20 CIP

A catalog record for this book is available from the British Library

ISBN 0-521-44500-0 Hardback
ISBN 0-521-44950-2 Paperback

Transferred to digital printing 2003

Contents

1

Derivation of a rate equation

Enzymes do not make reactions take place, they stimulate the rate at which reactions do take place. Any chemical reaction which proceeds in the presence of an enzyme will also proceed in the absence of the enzyme but at a *much* slower rate. Enzymes catalyze the rate of chemical reactions by lowering the activation energy of the reaction, and they do this in a manner which is highly specific for the reactants of the reaction. It was realized very early in the study of enzyme action that meaningful studies of enzyme action would, of necessity, involve the study of the kinetic behavior of the chemical reaction in the presence of the appropriate enzyme. It is still true that if one understands the kinetic behavior of the enzyme-catalyzed reaction, one also understands much about the mechanism of the enzymic reaction. This requires the investigation of the kinetic behavior of the enzymic reaction under conditions which are defined meticulously. Within the framework of this text, this will imply under steady state conditions. *Steady state*, as it applies to enzyme kinetics will be defined in this chapter and in chapter 2.

1.1 The role of 'diastase' in the early development of a theory

The enzyme-catalyzed hydrolysis of sucrose played an important role in the early development of a suitable equation to explain the kinetic behavior of enzyme-catalyzed reactions. One reason for the importance of this reaction was that the enzyme invertase was available in a reasonably pure form by the end of the nineteenth century when the principles of enzyme kinetics were established. In some of the early literature, this enzyme was called diastase. In fact, in some of the early literature all enzymes were called diastase. A second reason for the importance of sucrose hydrolysis in the development of enzyme kinetics was that the characteristics of acid-

catalyzed hydrolysis of sucrose had been well established by the latter part of the nineteenth century, and this allowed comparison of the acid-catalyzed hydrolysis with the enzyme-catalyzed reaction.

The hydrolysis of sucrose is the following reaction

$$\text{sucrose} + H_2O \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \text{glucose} + \text{fructose}$$

In the foregoing expression, k_1 and k_{-1} are second order rate constants, i.e. the rate of the reaction is proportional to the concentration of two reactants. If the reaction were carried out in an aqueous solution where the concentration of water would be approximately 55 M and if the concentration of sucrose were 1 M or less, the concentration of water would not change appreciably during the course of the reaction. Since the concentration of water would not change significantly even if the reaction continued to completion, one can assume $k_1' = k_1(H_2O)$, where k_1' is a pseudo-first order rate constant, the rate is proportional to the concentration of one reactant. The differential equation for the disappearance of sucrose with respect to time is

$$-\frac{d(A)}{dt} = k_1'(A) - k_{-1}(P)(Q) \tag{1.1}$$

where A = sucrose, P = glucose and Q = fructose. Throughout this book it will be assumed that the activity coefficient of any reactant is unity, thus the terms *concentration* and *activity* will be assumed to be interchangeable. If the concentration of either product were equal to zero or if k_{-1} were equal to zero, the second term on the right-hand side of eq. (1.1) would be equal to zero, and eq. 1.1 would become

$$-\frac{d(A)}{dt} = k_1'(A) \tag{1.2}$$

Equation (1.2) describes a reaction which would exhibit first order kinetic behavior. A plot of the rate of disappearance of A against the concentration of A should be a straight line which should pass through the origin with a slope equal to k_1'. This is shown in Figure 1.1.

Equation (1.2) can be rearranged and expressed in integral form.

$$\int_{A_0}^{(A)} \frac{d(A)}{(A)} = -k_1' \int_0^t dt \tag{1.3}$$

The result of the integration gives,

$$\ln(A) = -k_1' t + \ln A_0 \tag{1.4}$$

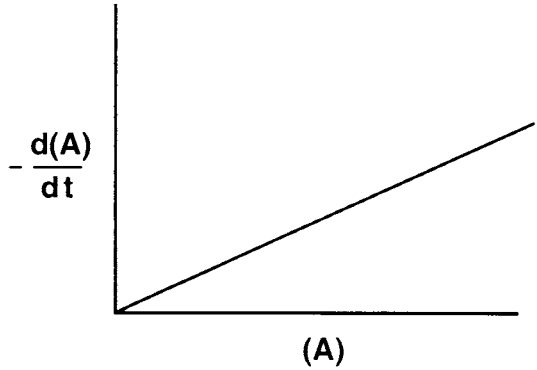

Fig. 1.1. Plot of the rate of disappearance of substrate A as a function of the concentration of A for a first order reaction.

where A_0 is the initial concentration of A.

$$(A) = A_0 e^{-k_1' t} \tag{1.5}$$

Substitution of eq. (1.5) into eq. (1.2) gives,

$$-\frac{d(A)}{dt} = k_1' A_0 e^{-k_1' t} \tag{1.6}$$

Plotting the rate of disappearance of A against time gives an exponential curve as is shown in Figure 1.2.

Investigations of the acid-catalyzed hydrolysis of sucrose were consistent with the hypothesis that the reaction followed first order kinetics. However,

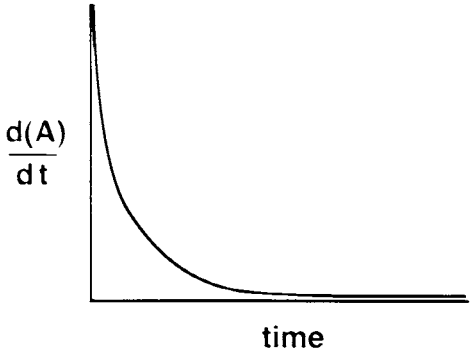

Fig. 1.2. Plot of the rate of disappearance of substrate A as a function of time for a first order reaction.

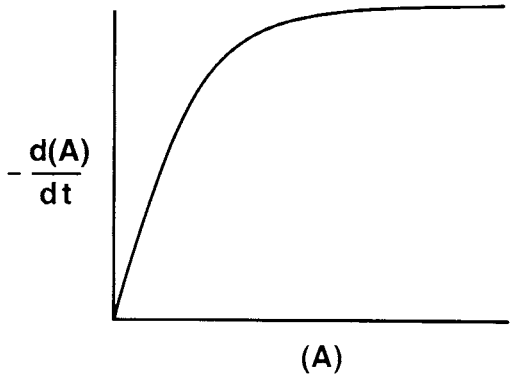

Fig. 1.3. Plot of the rate of disappearance of substrate A as a function of the concentration of A for a typical enzyme-catalyzed reaction.

investigations of the enzyme-catalyzed reaction led to observations which were perplexing at that time. Data which were obtained in experiments at low concentrations of substrate indicated first order kinetic behavior while experiments at high substrate concentrations suggested zero order kinetic behavior. That is, the reaction rate was a constant independent of substrate concentration. A careful analysis of the various results indicated that a plot of the rate of sucrose disappearance against sucrose concentration had the appearance shown in Figure 1.3. Numerous hypotheses were advanced to explain the kinetic behavior of enzyme-catalyzed processes[1], but none received widespread acceptance until the proposal suggested by Brown[2]. Brown's hypothesis was influenced by observations made by others. Wirtz[3] had reported that the proteolytic enzyme papain formed an insoluble complex with the substrate fibrin. This indicated that enzymes could combine with their substrates, but it did not provide evidence that the resulting complex was an obligatory intermediate in the reaction sequence. Additionally, O'Sullivan and Tompson[4] observed that invertase could tolerate a higher temperature in the presence of its substrate than in the absence of substrate. This observation was consistent with the hypothesis that sucrose could combine with invertase to form a complex which was more resistant to heat inactivation than was the native enzyme. Once again, this did not mandate that the enzyme-substrate complex was an obligatory intermediate in the reaction sequence. Finally, Emil Fischer's[5] "lock and key" explanation for enzyme specificity was best interpreted, at that time, in terms of an enzyme-substrate complex which is an obligatory intermediate in the reaction sequence. Thus, Brown suggested the following model

for enzyme-catalyzed reactions.

$$E + A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} EA \overset{k_2}{\rightarrow} E + P$$

In the foregoing reaction sequence, E represents the free enzyme while EA represents the complex of enzyme with substrate A. In this model, and throughout this text, the letters A, B, C, and D represent substrates while the letters P, Q, R and S represent products of the enzymic reaction. The forgoing model predicts that the reaction rate, i.e the increase of P with time should be $v = k_2(EA)$. Thus the rate of the reaction is proportional to the concentration of the EA complex. If the concentration of the enzyme were held constant and assays were run at increasing concentrations of substrate A, one would expect the concentration of the EA complex to be proportional to the concentration of A at low concentrations of A. Under those conditions, the kinetic behavior of the reaction would approximate first order kinetics. Inspection of Figure 1.3 shows that, at low concentrations of substrate A, the rate of the reaction is approximately a linear function of substrate concentration. On the other hand, if the concentration of the substrate were so high that essentially all the enzyme was present in the form of the EA complex, the rate of the reaction would be determined by the rate of decomposition of the EA complex to form a free enzyme and the product. At that point, increasing the concentration of the substrate would have no further effect on the rate of the reaction, and the reaction would exhibit zero order kinetic behavior with respect to substrate concentration. Inspection of Figure 1.3 shows that, at the highest concentrations of substrate, the plot of reaction rate versus the concentration of substrate is approximately a straight line with slope equal to zero. At intermediate concentrations of substrate, the curve represents a transition from first order to zero order kinetic behavior. The complete plot of reaction rate versus substrate concentration (substrate-saturation curve) is that of a rectangular hyperbola.

1.2 The basic assumptions on which derivation of an equation is based

The model which Brown proposed for enzymic reactions has withstood the test of time, but it is strictly intuitive and lacks a mathematical foundation. A mathematical treatment of this model was advanced first by a brilliant French scientist, Victor Henri[6]. It was Henri who derived the equation which is often attributed to Michaelis and Menten. Indeed, Michaelis and Menten[7] acknowledged that the purpose of their work was to provide

experimental affirmation of the mathematical formulation published by Henri. Based on the model proposed by Brown, one can write differential equations for the change in concentrations of each of the two enzyme species with respect to time.

$$\frac{d(E)}{dt} = -k_1(E)(A) + (k_{-1} + k_2)(EA) \tag{1.7}$$

$$\frac{d(EA)}{dt} = k_1(E)(A) - (k_{-1} + k_2)(EA) \tag{1.8}$$

As pointed out earlier,

$$v = -\frac{d(A)}{dt} = k_2(EA) \tag{1.9}$$

One might think that a mathematical expression for the rate of an enzymic reaction could be obtained by an analytical solution of the system of differential equations expressed in eqs. (1.7) and (1.8) and substitution of the expression for (EA) into eq. (1.9). Unfortunately, there is no analytical solution of eqs. (1.7) and (1.8). However, Henri reasoned that within a few milliseconds after the mixing of the enzyme with its substrate the concentrations of free enzyme and enzyme-substrate complex would become time-invariant. That is, for a given concentration of enzyme the relative amount of free enzyme and enzyme-substrate complex would be a function of substrate concentration, but the actual amount of each enzyme species would remain constant after the first few milliseconds. This assumption allows the differential equations of eqs. (1.7) and (1.8) to be replaced by the following linear algebraic equations.

$$-k_1(E)(A) + (k_{-1} + k_2)(EA) = 0 \tag{1.10}$$

$$k_1(E)(A) - (k_{-1} + k_2)(EA) = 0 \tag{1.11}$$

The foregoing two equations contain two unknown quantities, namely (E) and (EA), but it is not possible to solve the unknown quantities because the equations are not independent; in fact for the model under consideration they are identical. In order to derive an equation for the rate of an enzyme-catalyzed reaction, it is necessary to make a number of assumptions. These are,

$$E_t = (E) + (EA) \tag{1.12}$$

$$A_t \gg E_t \tag{1.13}$$

$$\frac{d(E)}{dt} = \frac{d(EA)}{dt} = 0 \tag{1.14}$$

$$(P) = 0 \tag{1.15}$$

In eq. (1.13), A_t is the total substrate concentration. The first three of these assumptions are essential for the derivation of the rate equation, the fourth assumption is made at this point as a matter of convenience and in chapter 4 the restriction imposed by eq. (1.15) will be removed. It is imperative that the reason for and the implications and validity of the assumptions expressed in eqs. (1.12) through (1.14) be understood. The logic behind the assumption expressed in eq. (1.12) is obvious. One could not conduct a valid assay if the total activity of the enzyme were changing during the assay. This assumption is often termed the *enzyme conservation expression.* However, this equation is indispensable mathematically for it provides a third equation and, therefore, a total of two independent equations which can be solved for the two unknown quantities. The reason for the remaining assumptions will be discussed in subsequent chapters.

1.3 The Briggs-Haldane steady state treatment of enzyme kinetic behavior

The derivation which will be presented is neither that of Henri nor that of Michaelis and Menten, but rather, the derivation of Briggs and Haldane[8,9]. The reason for following the Briggs-Haldane derivation is that it is a more general treatment. As noted earlier, the rate of the enzyme-catalyzed reaction for the model under consideration is $v = k_2(EA)$. The concentration of the free enzyme can be obtained from either eq. (1.10) or eq. (1.11).

$$(E) = \frac{(k_{-1} + k_2)}{k_1(A)}(EA) \tag{1.16}$$

Equation (1.16) can be substituted into eq. (1.12) and rearranged as

$$(EA) = \frac{k_1 E_t(A)}{k_{-1} + k_2 + k_1(A)} \tag{1.17}$$

The rate of the reaction is obtained by multiplying eq. (1.17) by k_2.

$$v = \frac{k_1 k_2 E_t(A)}{k_{-1} + k_2 + k_1(A)} \tag{1.18}$$

Equation (1.18) is identified as the Briggs-Haldane equation, the Michaelis-Menten equation and the Henri equation. Traditionally it is called the Michaelis equation and, reluctantly, that tradition will be followed in this book.

Equation (1.18) expresses the rate equation in terms of rate constants for the individual reactions. Throughout this book an enzymic rate equation expressed in terms of rate constants will be called the *rate equation* in the coefficient form. While the rate equation is usually derived in this form, it is not a useful form of the rate equation because most of the rate constants are generally inaccessible in investigations of the steady state behavior of enzymes. For this reason it is necessary to reformulate eq. (1.18) such that it is expressed in terms of parameters which can be determined in steady state studies. Throughout this book, these reformulations will be conducted in a similar manner. Equation (1.18) can be re-written as

$$v = \frac{\text{num. 1(A)}}{\text{constant} + \text{coef. A(A)}}, \tag{1.19}$$

where num. $1 = k_1 k_2 E_t$, constant $= k_{-1} + k_2$, and coef. $A = k_1$. The equation is reformulated by dividing both the numerator and denominator of the right hand side of eq. (1.19) by coef. A. The result is,

$$v = \frac{\frac{\text{num. 1}}{\text{coef. A}}(A)}{\frac{\text{constant}}{\text{coef. A}} + (A)} \tag{1.20}$$

The coefficient of the numerator term in eq. (1.20) is a constant, and the first term in the denominator of eq. (1.20) is also a constant. The equation is reformulated as,

$$v = \frac{V_{max}(A)}{K_m + (A)}, \tag{1.21}$$

where $V_{max} = k_2 E_t$ and $K_m = (k_{-1} + k_2)/k_1$. The Michaelis constant is K_m, and V_{max} is the maximal velocity. More precisely, V_{max} is the velocity of the reaction when the enzyme is saturated with the substrate. Throughout this book an enzymic rate equation expressed in terms of the steady state parameters will be called a *rate equation* in the kinetic form. Later in this chapter methods which provide for estimation of the steady state parameters, V_{max} and K_m will be discussed. Equations (1.18) and (1.21) both describe a rectangular hyperbola. Stated in more descriptive terms, they

are 1:1 order rational polynomials. A rational polynomial is a ratio of polynomials. A 1:1 order rational polynomial contains the independent variable, (A) in this case, to the first power in both the numerator and denominator.

It is informative to divide both the numerator and denominator of the right hand side of eq. (1.21) by the concentration of A,

$$v = \frac{V_{max}}{1 + \frac{K_m}{(A)}} \tag{1.22}$$

If the concentration of A were much less than K_m, such that $1 \ll K_m/(A)$, eq. (1.22) would become

$$v = \frac{V_{max}}{K_m}(A). \tag{1.23}$$

This equation describes a reaction which exhibits first order kinetic behavior, and the apparent first order rate constant is V_{max}/K_m. However, if the concentration of substrate were so great that $K_m/(A) \simeq 0$, eq. (1.22) would become

$$v \simeq V_{max}. \tag{1.24}$$

At this point the rate of the reaction would be independent of (A) and the reaction would exhibit zero order kinetic behavior with respect to substrate concentration. The mathematical definition of saturation of the enzyme with substrate A is, $K_m/(A) \simeq 0$. Finally, if $K_m = (A)$, $v = 1/2(V_{max})$. Thus the K_m is the concentration of the substrate which results in half maximal velocity, and the Michaelis constant is expressed in molarity.

The derivation presented here is that of Briggs and Haldane, and it differs from that developed by Henri and also that employed by Michaelis and Menten. In the case of the model under consideration, k_2 is the rate constant which includes the step which usually involves either the cleavage or formation of a covalent bond. If this step were very much slower than the other steps in the model, an equilibrium would be established between the free enzyme and the substrate and the enzyme-substrate complex. If such were the case, k_2 would be much smaller than k_{-1} and the Michaelis constant would be $K_m \simeq k_{-1}/k_1$. Thus, in the Briggs-Haldane treatment the Michaelis constant is a kinetic constant while in the Henri treatment it is a dissociation constant and, therefore, a thermodynamic constant. The matter of whether or not the reaction involving the cleavage of formation of

a covalent bond is very much slower than other steps in the reaction sequence will be discussed in chapter 9 of this book.

1.4 Estimation of steady state parameters

If one were to measure the velocity of an enzyme-catalyzed reaction in a series of assays in which the substrate concentration in each assay varied from one which was sufficiently small to result in a low rate relative to V_{max} to one where the substrate concentration were large enough to result in maximal velocity, one could plot the data and estimate both V_{max} and K_m. However, if, for example, the solubility of the substrate were limited in an aqueous solution, it might be impossible to estimate V_{max} and therefore K_m could also not be estimated. For this reason, efforts were made to rearrange eq. (1.18) in a linear form so that V_{max} could be obtained by extrapolation. Haldane and Stern[10], following the suggestions of B. Woolf, rearranged eq. (1.21) by dividing both sides of the equation by (A) and then inverting both sides of the equation to obtain

$$\frac{(A)}{v} = \frac{1}{V_{max}}(A) + \frac{K_m}{V_{max}}. \tag{1.25}$$

Equation (1.25) describes a linear relationship if $(A)/v$ were plotted against (A). The slope of the line is the reciprocal of V_{max} and the intercept of the $(A)/v$ axis is K_m/V_{max}. This same rearrangement was proposed by Hanes[11]. Haldane and Stern also noted that multiplying both sides of eq. (1.21) by $[K_m + (A)]$ and the rearrangement gives,

$$v = -K_m \frac{v}{(A)} + V_{max}. \tag{1.26}$$

Equation (1.26) is a linear relationship whose slope is $-K_m$ and whose intercept of the v axis is V_{max}. Lineweaver and Burk[12] utilized yet another rearrangement to obtain a linear form of eq. (1.21). This was accomplished by simply inverting both sides of eq. (1.21).

$$\frac{1}{v} = \frac{K_m}{V_{max}}\frac{1}{(A)} + \frac{1}{V_{max}}. \tag{1.27}$$

The resulting equation is that of a straight line whose slope is K_m/V_{max} and whose intercept of the $1/v$ axis is $1/V_{max}$. The point of intersection of the $1/(A)$ axis is $-1/K_m$.

There are additional linear forms of eq. (1.21), for example, the direct linear plot of Eisenthal and Cornish-Bowden[13], but the foregoing are the most widely used. It is important to recognize that these equations do not give rise to estimates of the steady state kinetic parameters with equal degrees of precision. Note that $1/v$, the dependent variable in the Lineweaver-Burk equation [eq. (1.27)] approaches infinity as 1/(A), the independent variable, approaches infinity. Hence, the Lineweaver-Burk plot places maximum weight on those observations which are made at low concentrations of the substrate, and those are the values which are associated normally with the largest experimental error. The converse is true of eq. (1.25) while eq. (1.26) places uniform weight on observations throughout the substrate-saturation curve. The significance is that the Lineweaver-Burk plot is the least desirable method for obtaining quantitative estimates of the steady state parameters. The objection to the use of the Lineweaver-Burk plot can be alleviated to some degree by utilizing a statistical program which employs weighting factors, but this necessitates the selection of an appropriate weighting factor. It is important to realize that the foregoing objection applies to the use of the Lineweaver-Burk plot as a means of obtaining quantitative estimates of K_m and V_{max} only, it does not argue against plotting data as a double reciprocal plot utilizing the estimates of the parameters which have been obtained by a more satisfactory method. Probably the most feasible method of obtaining quantitative estimates of the steady state parameters is the nonlinear regression method of Wilkinson[14]. The original publication outlines the method clearly for use with a calculator, but the procedure was outlined so well in Wilkinson's publication that it is easy to adapt it to a computer program to be run on a personal computer or even a programmable hand held calculator. This procedure is so elegant that there is little reason to obtain estimates of V_{max} and K_m by any other method provided the substrate-saturation curve is a rectangular hyperbola and the procedure is outlined in section 1.A.3 of the appendix to this chapter.

1.5 Problems for chapter 1

1.1 Derive expressions for the fraction of the total enzyme present as the free enzyme and for the fraction of the total enzyme present as the EA complex for the enzyme model considered in this chapter.

1.2 On the same sheet of graph paper, plot the $(E)/E_t$ and $(EA)/E_t$ as a function of $(A)/K_m$. Vary $(A)/K_m = 0.1$ to 10.

1.3 The following data were obtained in a substrate-saturation experiment.

(A)	v
mM	μmoles/minute
0.075	0.0120
0.100	0.0152
0.150	0.0205
0.200	0.0245
0.250	0.0280

Estimate K_m and V_{max} from a plot of $1/v$ versus $1/(A)$, and from a plot of $(A)/v$ versus (A), and finally from a plot of v versus $v/(A)$.

Appendix: A brief look at statistical analysis

1.A.1 Definition of a few statistical terms

It is the purpose of this appendix to provide a brief account of the simpler statistical analyses employed in enzyme kinetics. The first statistic is the arithmetic mean or average. If one were to measure the change in absorbance at 340 mμ in a cuvette in a given time interval after a dehydrogenase had been added to a reaction medium containing NAD^+ and the appropriate oxidizable substrate several times, one would record a number of slightly different values. If the several values were designated $\sum Y_i$ and n were the number of observations, the average change in absorbance would be

$$\bar{Y} = \frac{\sum Y_i}{n} \tag{1.A.1}$$

The arithmetic mean does not give any indication of the amount of scatter in the observations. A measure of the accuracy of the mean should be related to the deviations about the mean, but, in theory, the sum of the deviations greater than the mean should be offset by the deviations less than the mean. Hence the sum of the deviations should be equal to zero. For this reason, and for theoretical reasons that will not be discussed here[15], the deviations about the mean are squared. The variance is defined as the sum of the squares of the deviations divided by the degrees of freedom. If there is one parameter measured, the degrees of freedom is given by $n - 1$. Therefore the expression for the variance is,

$$\text{variance} = s^2 = \frac{\sum (Y - \bar{Y})^2}{n - 1} \tag{1.A.2}$$

The squared term in the numerator can be expanded as

$$\sum (Y^2 - 2Y\bar{Y} + \bar{Y}^2) = \sum Y^2 - 2\sum Y \sum \frac{Y}{n} + \left(\frac{\sum Y}{n}\right)^2 n$$

The variance can be expressed as follows:

$$s^2 = \frac{\sum Y^2 - \dfrac{(\sum Y)^2}{n}}{n - 1} \tag{1.A.3}$$

The standard deviation is defined as the square root of the variance.

$$\text{s.d.} = \sqrt{\frac{\sum Y^2 - \frac{(\sum Y)^2}{n}}{n-1}} \tag{1.A.4}$$

Standard deviation and standard error are similar terms. They may be used interchangeably if the statistic to which they apply is specified. The coefficient of variation is defined as follows:

$$\text{c.v.} = \frac{\sqrt{\text{variance}}}{\bar{Y}} \times 100 = \frac{S}{\bar{Y}} \times 100 \tag{1.A.5}$$

The symbol s^2 is defined as the variance of a given sample while σ^2 is defined as the variance of the population from which the sample is taken. The former is estimated from the data, but the statistician is usually interested in the variance of the population rather than that of the sample of the population. In the same manner, the arithmetic mean of the sample is defined as $\bar{Y}$ while the arithmetic mean of the population is μ.

1.A.2 Linear regression

A procedure which is employed extensively in the analysis of enzyme kinetic data, as well as in all of biochemistry, is linear regression. Linear relationships are well understood in mathematical terms. Three linear transformations of the Michaelis equation have been presented in this chapter. In general terms, the assumption of a linear relationship implies that a dependent variable, Y, is a linear function of an independent variable X. In the case of the Lineweaver-Burk plot, $1/v$ is the dependent variable while 1/(A) is the independent variable. In the case of the (A)/v versus (A) plot, (A)/v is the dependent and (A) is the independent variable. However, the linear regression, as here presented, is based on four assumptions. These assumptions are,

1) It is assumed that the independent variable, X, is measured without error.
2) The expected value of the dependent variable for a given value of the independent variable is,

$$Y = \text{A} + \text{B}X. \tag{1.A.6}$$

3) For any given value of X, the observed Y values are distributed independently and normally. This is represented by,

$$Y_i = \text{A} + \text{B}X + \varepsilon_i. \tag{1.A.7}$$

where ε_i is the error in the estimate.
4) It is assumed that the variance around the regression line is constant and, therefore, independent of the magnitude of X or Y.

The aim in the linear regression is to calculate the values of A and B in eq. (1.A.7) such that Y can be estimated for any given value of X. Thus, the difference between the observed values of Y_i and the value estimated from the regression line would be

$$r_i = \hat{Y}_i - Y_i \tag{1.A.8}$$

where r_i is the residual and $\hat{Y}_i$ is the estimated value of Y_i. As with variance, the residuals are squared.

$$\sum r_i^2 = \sum(\hat{Y}_i - Y_i)^2 = \sum(\mathrm{A} + \mathrm{B}X_i - Y_i)^2 \tag{1.A.9}$$

The problem is to find the values for A and B which minimize $\sum r_i^2$. This is accomplished by setting the following derivatives equal to zero.

$$\frac{\mathrm{d}\sum(\mathrm{A} + \mathrm{B}X - Y)^2}{\mathrm{dA}} = 0, \quad \frac{\mathrm{d}\sum(\mathrm{A} + \mathrm{B}X - Y)^2}{\mathrm{dB}} = 0$$

An expansion of the squared numerator and differentiation gives the following expressions.

$$\frac{\mathrm{d}\sum r^2}{\mathrm{dA}} = 2\sum \mathrm{A} + 2\mathrm{B}\sum X - 2\sum Y = 0$$

$$\frac{\mathrm{d}\sum r^2}{\mathrm{d}B} = 2\mathrm{A}\sum X + 2\mathrm{B}\sum X^2 - 2\sum XY = 0$$

The following equations arise from the foregoing

$$\sum Y = n\mathrm{A} + \mathrm{B}\sum X \tag{1.A.10}$$

$$\sum XY = \mathrm{A}\sum X + \mathrm{B}\sum X^2 \tag{1.A.11}$$

Equations (1.A.10) and (1.A.11) can be solved for A and B. Note that A is the point of intersection of the regression line with the Y axis and B is the slope of the regression line, but the latter is more often referred to as the *regression coefficient*.

A variety of computer-based library programs which provide for numerical solution of eqs. (1.A.10) and (1.A.11) as well as providing estimates of the standard error for both parameters are available. There are also books which contain the source code for such programs[16]. Nevertheless, it seems appropriate to provide a brief account here of how such computations are conducted[17]. Equations (1.A.10) and (1.A.11) can be generalized as,

$$a_{11}x_1 + a_{12}x_2 = b_1$$

$$a_{21}x_1 + a_{22}x_2 = b_2$$

These equations can be written in matrix form

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix} \begin{vmatrix} x_1 \\ x_2 \end{vmatrix} = \begin{vmatrix} b_1 \\ b_2 \end{vmatrix}$$

An augmented matrix which consists of the A matrix, the B vector from the foregoing and an identity matrix can be constructed

$$|\mathrm{A} \quad \mathrm{B} \quad \mathrm{I}| = \begin{vmatrix} a_{11} & a_{12} & b_1 & 1 & 0 \\ a_{21} & a_{22} & b_2 & 0 & 1 \end{vmatrix}$$

After the equations have been solved, the resultant augmented matrix will be

$$|\mathrm{I} \quad x \quad \mathrm{A}^{-1}| = \begin{vmatrix} 1 & 0 & x_1 & c_{11} & c_{12} \\ 0 & 1 & x_2 & c_{21} & c_{22} \end{vmatrix}$$

In the foregoing, the elements c_{ij} are the elements in matrix A^{-1}. The inversion of matrix A is accomplished by two types of operations. The first of these is

normalization in which all of the elements in a row of the original augmented matrix are divided by the first non-zero element of the row. By a repetition of this operation, the diagonal elements of A are converted to ones. The second operation is *reduction* in which the non-diagonal elements of A are converted to zeros. Normalization of the first row of the original augmented matrix gives

$$\begin{vmatrix} 1 & \frac{a_{12}}{a_{11}} & \frac{b_1}{a_{11}} & \frac{1}{a_{11}} & 0 \\ a_{21} & a_{22} & b_2 & 0 & 1 \end{vmatrix}$$

Reduction is performed by multiplying each element of the normalized first row by a_{21} and subtracting the product from the corresponding element in row 2. The result of this reduction is

$$\begin{vmatrix} 1 & \frac{a_{12}}{a_{11}} & \frac{b_1}{a_{11}} & \frac{1}{a_{11}} & 0 \\ 0 & \frac{a_{11}a_{22}-a_{12}a_{21}}{a_{11}} & -\frac{a_{11}b_2-a_{21}b_1}{a_{11}} & -\frac{a_{21}}{a_{11}} & 1 \end{vmatrix}$$

Normalization of the second row gives

$$\begin{vmatrix} 1 & \frac{a_{12}}{a_{11}} & \frac{b_1}{a_{11}} & \frac{1}{a_{11}} & 0 \\ 0 & 1 & \frac{a_{11}b_2-a_{21}b_1}{a_{11}a_{22}-a_{12}a_{21}} & -\frac{a_{21}}{a_{11}a_{22}-a_{12}a_{21}} & \frac{a_{11}}{a_{11}a_{22}-a_{12}a_{21}} \end{vmatrix}$$

The final reduction is accomplished by multiplying each element of the normalized second row by a_{12}/a_{11} and subtracting the product from the corresponding element of row 1. The result is the final augmented matrix.

$$\begin{vmatrix} 1 & 0 & \frac{a_{22}b_1-a_{21}b_2}{a_{11}a_{22}-a_{12}a_{21}} & \frac{a_{22}}{a_{11}a_{22}-a_{12}a_{21}} & -\frac{a_{21}}{a_{11}a_{22}-a_{12}a_{21}} \\ 0 & 1 & \frac{a_{11}b_2-a_{21}b_1}{a_{11}a_{22}-a_{12}a_{21}} & -\frac{a_{21}}{a_{11}a_{22}-a_{12}a_{21}} & \frac{a_{11}}{a_{11}a_{22}-a_{12}a_{21}} \end{vmatrix}$$

The symbols from eqs. (1.A.10) and (1.A.11) can be substituted into the foregoing matrix to give

$$\begin{vmatrix} 1 & 0 & \frac{\sum X^2 \sum Y - \sum X \sum XY}{n\sum X^2 - (\sum X)^2} & \frac{\sum X^2}{n\sum X^2 - (\sum X)^2} & -\frac{\sum X}{n\sum X^2 - (\sum X)^2} \\ 0 & 1 & \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} & -\frac{\sum X}{n\sum X^2 - (\sum X)^2} & \frac{n}{n\sum X^2 - (\sum X)^2} \end{vmatrix}$$

It can be seen from the foregoing that A in eqs. (1.A.10) and (1.A.11) is equal to the first element in the third column of resultant augmented matrix, and B is equal

to the second element in the third column of the matrix. Thus,

$$A = \frac{\sum X^2 \sum Y - \sum X \sum XY}{n\sum X^2 - (\sum X)^2} = \frac{(\sum X^2 \sum Y - \sum X \sum XY)/n}{\sum X^2 - \frac{(\sum X)^2}{n}} \tag{1.A.12}$$

$$B = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{\sum XY - \sum X \frac{\sum Y}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}} \tag{1.A.13}$$

The variance about the regression line is estimated by

$$s^2 = \frac{\sum (y - A - BX)^2}{n-2} = \frac{\sum Y^2 - A\sum Y - B\sum XY}{n-2} \tag{1.A.14}$$

The degree of freedom in eq. (1.A.14) is $n-2$ because both A and B are estimated from the observed data. Without providing a derivation, the standard error on the estimate of A is

$$\text{s.e.}_A = \sqrt{s^2 \times \frac{\sum X^2}{n\sum X^2 - (\sum X)^2}} \tag{1.A.15}$$

It should be noted that the standard error for A is equal to the square root of the variance times the first element of the fourth column of the augmented matrix. In like manner, the standard error for the estimation of B, the regression coefficient, is

$$\text{s.e.}_B = \sqrt{s^2 \times \frac{n}{n\sum X^2 - (\sum X)^2}} \tag{1.A.16}$$

The standard error of the regression coefficient is equal to the square root of the variance times the second element of the last column in the augmented matrix. Thus the diagonal elements in the inverse matrix A^{-1} are factors used in estimation of the standard errors. The variables which are required for estimation of A and B and their standard errors are $n, \sum Y, \sum Y^2, \sum X, \sum X^2, \sum XY$. The equations presented in this text may appear to differ from those found in reference books on statistical analysis[15], but they are equivalent. Statisticians rearrange the equations to improve the computational efficiency. The purpose here has been to outline the derivations in a fairly straightforward manner.

1.A.3 Non-linear regression in enzyme kinetic analysis

The non-linear regression method proposed by Wilkinson for estimation of the steady state enzyme kinetic parameters will be outlined here in essentially the manner it has been presented in the original publication[14]. Non-linear regression requires a preliminary estimate of the parameters, and these are obtained by linear regression similar to that described in the previous discussion. However, Wilkinson employed a weighted linear regression of (A)/v versus (A). The following

expressions provide for the preliminary estimates of the parameters.

$$|D| = \sum v^3 \sum \frac{v^4}{(A)^2} - \sum \frac{v^3}{(A)} \sum \frac{v^4}{(A)} \tag{1.A.17}$$

$$K_m = \left[\sum v^4 \sum \frac{v^3}{(A)} - \sum v^3 \sum \frac{v^4}{(A)} \right] \Big/ |D| \tag{1.A.18}$$

$$V_{max} = \left[\sum v^4 \sum \frac{v^4}{(A)^2} - \left(\sum \frac{v^4}{(A)} \right)^2 \right] \Big/ |D| \tag{1.A.19}$$

The non-linear aspect of the Wilkinson method is based on the assumption that if a function is non-linear in a parameter, c, the following linear approximation may be used.

$$f_{v,c} \cong f_{v,c^0} + (c - c^0) f'_{v,c^0} \tag{1.A.20}$$

where c^0 is a provisional estimate of c and f' is the first derivative of f with respect to c. In terms of enzyme kinetics, eq. (1.A.20) becomes

$$V \cong \frac{V_{max}}{V^0_{max}} \left[\frac{V^0_{max}(A)}{K^0_m + (A)} - (K_m - K^0_m) \frac{V^0_{max}}{(K^0_m + (A))^2} \right] \tag{1.A.21}$$

To initiate the calculations, the preliminary estimates of K_m and V_{max} from linear regression are used as the provisional estimates of the parameters. The following calculations lead to updated estimates of the parameters.

$$f = \frac{V^0_{max}(A)}{K^0_m + (A)} \tag{1.A.22}$$

$$f' = -\frac{V^0_{max}(A)}{(K^0_m + (A))^2} \tag{1.A.23}$$

$$|D| = \sum f^2 \sum f'^2 - (\sum f f')^2 \tag{1.A.24}$$

$$b_1 = [\sum f'^2 2 \sum vf - \sum f f'^2 \sum vf'] / |D| \tag{1.A.25}$$

$$b_2 = [\sum f^2 \sum vf' - \sum f f' \sum vf] / |D| \tag{1.A.26}$$

$$K_m = K^0_m + \frac{b_2}{b_1} \tag{1.A.27}$$

$$V_{max} = V^0_{max} \times b_1 \tag{1.A.28}$$

The updated estimate of K_m is tested against the provisional estimate. If the two are sufficiently close, for example, if $abs(K_m - K^0_m)/K^0_m \preceq 0.001$, the updated parameters are accepted as the best estimates. If on the other hand, the updated K_m does not pass the test, the provisional estimates of the parameters are replaced by the updated estimates and the calculations embodied in eqs. (1.A.22) through (1.A.28) are repeated. It should be understood that, as K^0_m approaches the best estimate, the value of b_1 approaches unity and b_2 approaches zero. After the best estimates have been obtained, the variance can be calculated as

$$s^2 = \frac{\sum v^2 - b_1 \sum vf - b_2 \sum vf'}{n - 2} \tag{1.A.29}$$